# *EXHIBIT 3*

**Report Regarding the Safety Evaluation of Gardasil**
**in Relations to Auto-Immune Diseases**

Martin Kulldorff, PhD

## 1. PERSONAL BACKGROUND

As a biostatistician and epidemiologist, I have done research on post-market vaccine safety evaluations for around two decades, primarily working with FDA and CDCs Vaccine Safety Datalink.

I have a bachelor's degree in mathematical statistics from Umeå University in Sweden (1984) and a Ph.D. in operations research with a focus on applied probability and statistics from Cornell University (1989). I have since worked as a junior faculty member in the Department of Statistics at Uppsala University in Sweden, as a government scientist in the Biometry Branch at the National Cancer Institute, as an associate professor in the Department of Community Medicine at the University of Connecticut Medical School and as an associate and full professor of medicine at Harvard Medical School. I have worked on many federal research grants, but I have never received funding from the pharmaceutical industry. Having recently left academia, I currently work as a biostatistical and epidemiological consultant, working for example with the New York City Department of Health and Mental Hygiene and Acumen LLC.

With colleagues, I have authored over 200 peer-reviewed articles, of which over 40 are vaccine related, concerning vaccines against for examples, measles, rotavirus, pertussis, human papilloma virus, pneumococcal disease, meningococcal disease, tetanus, diphtheria, varicella, herpes zoster, influenza, covid, and rabies. These peer-reviewed vaccine articles have been published in major journals such as American Journal of Epidemiology, American Journal of Preventive Medicine, New England Journal of Medicine, Pediatrics, Pharmacoepidemiology and Drug Safety, PLoS

Medicine, Proceedings of the National Academy of Sciences, and Vaccine. Most of the time we have not found any major problems with the vaccines studied, but there have been exceptions such as the discovery of an excess risk of seizures after the MMRV vaccine against measles, mumps, rubella and varicella (Klein et al., 2010)

I developed some of the methods that used by FDA and CDC in their routine post-market vaccine safety surveillance work, including sequential analysis methods for the rapid detection of potential safety problems using weekly or monthly data feeds (Kulldorff et al. 2011; Yih et al. 2011); self-control methods that can adjust for any between-person confounding (Klein et al., 2010; Yih et al 2018); and data mining methods to simultaneously evaluate thousands of potential safety problems while adjusting for the multiple testing inherent in such activities (Kulldorff et al. 2003; Li et al. 2018).

I have served on two CDC ACIP vaccine safety working groups, concerning the measles-mumps-rubella-varicella (MMRV) and Covid vaccines. I was terminated from the latter working group, when I objected to the CDC instituted "pause" on the Johnson and Johnson Covid vaccine in 2021, although CDC lifted the pause four days later, in accordance with my views. I served three years on FDA's Drug Safety and Risk Management Safety Committee on non-vaccine work.

## 2. STUDY EVALUATION CRITERIA

I have been asked for my expert evaluation of the following three areas:

1. The response by the Market Authorization Holder (MAH) the Article 20 Request by the European Medical Agency (EMA).
2. The VAERS data analysis of Gardasil and auto-immune diseases by Dr. Lucija Tomljenovic

3. Chao et al. Surveillance of autoimmune conditions following routine use of quadrivalent human papillomavirus vaccine, Journal of Internal Medicine, 2011.

In writing this report, I take the same scientific approach that I have used for decades in detection and evaluating potential vaccine safety problems. The quality of a medical research study depends on (i) the data set used, (ii) the epidemiological design, (iii) the statistical analysis, and (iv) the interpretation of the results. If there are multiple different analyses in an article, each must be evaluated separately in terms of the design and analysis. For the studies that I have been asked to evaluate, I comment on the quality of the data set used, the design of the study, and the statistical analysis. I conclude with comments on the authors' interpretation of their results and how I think the results should be interpreted.

Vaccines are important for public health, and for the sake of vaccine confidence, it is critical to (i) rapidly and thoroughly investigate any potential problem, (ii) use well designed studies with adequate sample size and statistical power so that a null result is evidence that the vaccine is safe for the outcomes studied, (iii) never dismiss a problem based on studies that are either badly designed or analyzed, or under powered. One must also (iv) honestly report the results, whether there is evidence for a problem, evidence that the vaccine is safe, or no evidence in either direction.

## 3. COMMENTS ON THE MAH RESPONSE TO EMA'S ARTICLE 20 REQUEST

Due to reports of patients being diagnosed with complex regional pain syndrome (CRPS) and postural orthostatic tachycardia syndrome (POTS) after receiving the Gardasil vaccine, the European Medical Agency (EMA) requested a response from Merck, the vaccines Marketing Authorization Holder (MAH). The EMA asked several very specific questions from their Pharmacovigilance Risk Assessment Committee

(PRAC), and Merck responded with a 188-page report. My assessment of the scientific adequacy of their report is below.

**PRAC Question # 1: Clinical Trial Data** (Section 1.1.1)

Clinical trials are by nature limited in sample size. Among 15,801 recipients of Gardasil 9, the MAH reports two suggestive and one definite case of postural orthostatic tachycardia syndrome (POTS), for an incidence of one case per 7,900 or 15,801 vaccinees respectively, with no POTS cases in the placebo group. For vaccines and a serious adverse reaction like POTS, these are high point estimates. Because of the small sample size, it may have been a chance finding, but it is incorrect and dishonest for the MAH to claim that *"the incidence of the cases suggestive of … POTS was found to be extremely low."* Moreover, with no cases in the placebo group, it is nonsense to claim that the incidence was *"similar between the 9vHPV … and placebo groups."* (p.5). It was also misleading for the MAH to state that *"this assessment does not suggest an association between HPV vaccination and … POTS" (p.5).* While possibly a chance occurrence, it was something that should have been urgently investigated using observational data with a larger sample size in a post-market setting. Vaccines are important for public health, and vaccine confidence requires an honest evaluation of all potential adverse reactions rather than misleading and dishonest dismissals of potential problems.

**PRAC Question # 1: Post Marketing Data** (Section 1.1.2)

The post-marketing section of the MAH response only includes spontaneous report data. Spontaneous report data are typically evaluated through proportional reporting ratios, reporting odds ratios, or similar analytical methods. Instead of such standard analyses, the MAH response devotes a large number of pages to the descriptive nature of the spontaneous reporting data. The only data analysis is an estimate of the incidence rate by comparing the reported cases with the estimated number of vaccinated patients, but it is a well-accepted principle within the scientific community that spontaneous report data by definition suffer from under reporting,

not capturing all cases. Such an "analysis" cannot and should not be used to dismiss a potential safety problem. This is exactly why proportional reporting ratios and similar methods are used. Without that, this 77-page section offers very little scientifically useful information.

While spontaneous reports are important to quickly identify potential safety problems, it is not the only data available to rigorously evaluate a potential problem. Rather than exclusively focusing on spontaneous reporting data, one would also have expected an analysis of electronic health records or insurance claims data, which Merck routinely utilizes for many other studies.

In summary, the spontaneous report data were not analyzed using the standard and best methods for such data, and no other post-marketing data sources were used.

**PRAC Question # 1: Literature Review** (Section 1.1.3)
The literature mainly consists of a series of case reports. While such reports are important for clinical understanding, they cannot rule out causal relationships.

**PRAC Question # 2** (Section 1.2)
The MAH has done a careful review of the CRPS and POTS cases that occurred as part of the clinical trials. As with the literature review in section 1.1.3, they neither confirm nor rule out a causal relationship.

**PRAC Question # 3** (Section 1.3)
The MAH was tasked to *"provide an analysis of the observed number of post-marketing cases of CRPS and POTS in association with their HPV vaccine in comparison to those expected in the target population."* An expected approach would have been to do a comparison using electronic health data or health insurance claims data, such as for example, the same Kaiser Permanente data set that they used for the Chao et al. study in 2011.

Instead, they used a mismatch of different data sets for which the numbers are neither reliable nor comparable. For the observed cases after vaccination, they used spontaneous reporting data with known under-reporting. Such data cannot reliably be compared to background rates in the target population, however carefully and accurately those background rates are calculated. They should instead be analyzed by looking at the total number of spontaneous reports received after Gardasil vaccination, to see what proportion of those are for CRPS/POTS. This proportion should then be compared with that proportion that is observed for other vaccines, using for example proportional reporting ratios or reporting odds ratios (Evans et al., 2001; Rothman et al., 2004).

As mentioned, even perfect background rates would not generate a proper comparison with spontaneous reporting data, but to make things worse, the backgrounds rates used are flawed. Specifically, there is a severe problem with how the MAH report calculated the expected number of cases. There are uncertainties both in the estimated denominator (number of people exposed to the vaccine, #1-#4 below) and in the numerator (background rate of the number of observed cases among the unexposed).

1. The number of distributed doses is typically accurate, so that part is reliable.
2. As acknowledged by the MAH, there are uncertainties in how many of the distributed doses are actually administered.
3. Even if we knew the number of administered doses, which we do not, there is additional uncertainty in the number of people exposed, since we do not know how many people received one, two or three doses.
4. While we roughly know the age distribution of those vaccinated, there is some uncertainty here as well.

For CRPS, the incidence rate per 100,000 person years were assumed to be 14.9 for girls ages 10 to 19 and 28.0 for women ages 20 to 29. From the main text, it is unclear where these numbers come from. It just says that *"a literature review was conducted to identify background incidence rates and prevalence of CRPS and POTS."* In the legend of Table 12 there is a reference to Appendix A for the estimated incidence rates. For CRPS, they reference the Sandroni et al. (2003) study from Olmsted County in Minnesota and the de Mos et al. (2007) study from the Netherlands, with the two studies having widely different incidence estimates. In Minnesota, the CRPS background incidence rate was 2.15 among girls ages 10 to 19 and 6.81 among women ages 20-29, but rather than taking a weighted average, the MAH used the higher incidence numbers of 14.9 and 28.0 from the Netherlands, not even mentioning the lower numbers from Minnesota. They then assume that incidence of CRPS is the same in all countries as in the Netherlands, but very few diseases have the same incidence rate everywhere.

The calculations of observed and expected cases for POTS are even worse. As stated in the report, *"there are minimal prevalence data for POTS in the literature, and to our knowledge, no incidence rates for POTS have been reported."* Instead, they estimate the background incidence rates of POTS using chronic fatigue syndrome (CFS), for which the published estimates vary more than a hundredfold for the same age- and gender groups. For example, for both adults of both genders, the point estimates per 100,000 person years are <1 in 4 US cities (Reyes, 1997), 9 in UK (Nacul 2011), 13.16 in Olmsted County (Vincent 2012), 55 UK (Donegan 2013), 180 in Kansas (Reyes 2003), and 370 in Scotland (Lawrie 1997).

The inaccurate nature of these analyses is reflected in Table 13. If the truth is that there is no relationship between Gardasil and CRPS, then the observed and expected numbers should be roughly equal, except for some randomness. For the worldwide numbers with a two-year follow-up in Table 13, there were 29 observed cases that met or partially met the case criteria among the vaccinated, compared to an expected

number somewhere between 161 and 9,932, depending on the assumptions made by the MAH.

By playing with assumptions and numbers in this way, the MAH report is not scientifically sound. With about the same amount of time and effort, MAH could have used electronic health records or insurance claims data to conduct a proper analysis to actually answer the question asked.

**PRAC Question # 4** (Section 1.4)

With limited sample size in most randomized vaccine trials, observational studies can be critical for evaluating potential adverse reactions. By recognizing that *"to our knowledge, there are no published studies evaluating the association between HPV vaccines and CRPS or POTS,"* the MAH is de facto acknowledging that they do not know whether the vaccines cause CRPS or POTS. They then proceed to discuss observational studies that looked at other auto-immune diseases, but just because the vaccine does not cause A, B, C or D, does not mean that it does not cause CRPS or POTS.

The MAH states that *"there is no epidemiologic evidence … for an association between HPV vaccine and CRPS (POTS)."* (p.168, 171); yet there were not proper studies at that time to evaluate this association from an epidemiological standpoint.

**PRAC Question # 5** (Section 1.5)

As the last question, the MAH was asked to *"discuss the need for possible risk minimization tools and provide proposals."* In light of the fact that they did not conduct a proper analysis as to whether there may be a relationship between Gardasil and CRPS and/or POTS, it is surprising that they do not propose to conduct a proper study to evaluate the question, using spontaneous reports, electronic health data, or health insurance claims.

The questions asked by the EMA regarding the Gardasil vaccine are appropriate and important. While EMA cannot control how Merck responds to their questions, the EMA should not have accepted this report as an adequate response to their queries. The response shows a shocking lack of ability by Merck to properly evaluate and discuss the safety of their own products.

## 4. COMMENTS ON THE VAERS DATA ANALYSIS OF GARDASIL AND AUTO-IMMUNE DISEASES BY DR. TOMLJENOVIC

Using spontaneous reporting data from VAERS, Dr. Lucija Tomljenovic has examined the relationship between Gardasil and various auto-immune disease outcomes, including postural orthostatic tachycardia syndrome (POTS) and primary ovarian insufficiency (POI) (Tomljenovic, 2024). Using reporting odds ratios, this was done using a proper and standard statistical analysis. In fact, she did one of the things that should have been done as part of the MAH report to the EMA.

**Data**

Administered jointly by FDA and CDC, the Vaccine Adverse Event Reporting System (VAERS) collects "spontaneous" reports from health professionals, patients and others concerning suspected adverse reactions after any vaccine. It is a critically important vaccine safety system, but the data also has weaknesses.

The main strengths of these data are that any suspected adverse reaction can be reported no matter where or when it occurs. Another important strength is that it does not depend on a pre-determined disease classification system that is used for electronic health and insurance claims data. Hence, it can detect unusual disease syndromes that do not fit perfectly within the established ICD-10 framework.

The main weaknesses of the data are that there is both under reporting and over reporting, as there is no well-defined population for which all health events of interest

are extracted. There is under reporting because not everyone with the adverse event of interest reported it to VAERS. There is over reporting because some VAERS reports are for adverse events not caused by the vaccine. Another weakness is that VAERS data is only tabulated in broad age groups rather than the exact age. This is unfortunate, as it makes it hard to properly adjust for age, which could bias the results of any analysis.

While we know how many reports there are for the adverse event of interest, we do not know how many people took the vaccine without having the adverse event. We just know how many people who took the vaccine reported a different adverse event. Hence, it would be hopeless to try and compare the observed numbers with population-based backgrounds rates.

With the lack of denominator data, VAERS data are typically analyzed based on a 2x2 table consisting of (i) the number of spontaneous reports of outcome X received after vaccine Y, comparing it with (ii) the number of other reports received after vaccine Y, (iii) the number of reports of outcome X received after other vaccines, and (iv) the number of other reports received after other vaccines. For example, if 10% of all VAERS reports after a measles vaccine is for febrile seizures, while only 2% of all VAERS reports after other childhood vaccines are for febrile seizures, that may be an indication that the measles vaccine causes febrile seizures. This is what Dr. Tomljenovic has done.

VAERS data can be used either (i) in a data mining exercise or (ii) to look at a small set of specific outcomes after a particular vaccine of interest. Dr. Tomljenovic appropriately used it for the second purpose, since she is evaluating pre-specified hypotheses about POTS and a few other outcomes. It would then be wrong to use a data mining method such as empirical Bayes Poisson shrinkage or a tree-based scan statistic. When hundreds or thousands of potential adverse events are simultaneously evaluated, then data mining methods appropriately adjust for the

multiple testing inherent in the many adverse events evaluated. But if only a few outcomes are of interest, then data mining methods over adjust, resulting in a loss of statistical power, artificially making it much harder to detect a real problem.

**Reporting Odds Ratios**

Dr. Tomljenovic has analyzed VAERS data for two primary time periods: 2006-2015 and 2006-2024. The latter is obviously what should be looked at today in determining if the vaccine may cause these adverse reactions and if further studies are warranted. The former time period is of interest in terms of what Merck would have had access to around the time of their MAH report to EMA, showing the analyses they could and should have performed as part of their vaccine safety evaluation efforts. Hence, while one would normally not do multiple analyses with different end dates, in this case both periods are of interest but for different reasons.

Two common and appropriate statistical methods for analyzing VAERs data are proportional reporting ratios and reporting odds ratios. While the exact interpretations of the point estimates are different, both methods are valid for analyzing the 2x2 tables generated from VAERS data, and the methods tend to produce similar results in terms of statistical significance.

Dr. Tomljenovic has calculated and presented reporting odds ratios. The statistical calculations that I checked were done correctly and results are logical in light of the 2x2 tables being analyzed. A few of the results are reproduced in Table 1 below.

**Results and Interpretation**

For girls ages 6-17, the 2006-2024 VAERS data show a clear and obvious signal for both postural orthostatic tachycardia syndrome (POTS) and for primary ovarian insufficiency (POI), with a reporting odds ratio of 11.1 and 10.0 respectively (Table 1). For both outcomes, $p<0.0001$ with the lower 95% confidence intervals above 7 and $p<0.0001$. These are very concerning numbers.

The analyses of women ages 18-29 and boys ages 6-17 give similar results. Since these groups are non-overlapping, these two analyses are statistically independent from the analysis of girls ages 6-17, providing independent evidence for a signal.

With such low p-values, high lower limits on the 95% confidence intervals, and consistent results across non-overlapping population groups, these signals are not random signals caused by chance. The nature of the data does not allow for age-adjustment, and there is likely some bias due to the inability to adjust for age, but age or other covariates are unlikely to explain more than at most a small fraction of the high ROR estimates.

With VAERS data, one must always be concerned with reporting bias. This is especially true after there is an initial suspicion of a safety problem. I was therefore pleased to see that Dr. Tomljenovic did secondary analyses for the shorter time periods from 2006-2009, 2006-2010 and 2006-2012, before there was much public concern about Gardasil and POTS. While the sample size is by default smaller, with wider confidence intervals, the results are similar pointing in the same direction. This makes reporting bias less of a concern than it otherwise would have been. Still, it is important to also conduct analyses using other observational data with denominators and unbiased case assessment between vaccinated and unvaccinated.

These two VAERS signals should not and cannot be ignored. The signals were already present in the VAERS data back in 2015 (Table 1), and an immediate and thorough evaluation was warranted at that time. It is hard to explain why such evaluations were not done by Merck a long time ago. There are no scientific or public health reasons.

While POTS and POI stand out in the ROR analyses of the VAERS data, there are also other auto-immune definitions and diseases that warrant further investigations.

| | | | Postural Orthostatic Tachycardia Syndrome (POTS) | | | Primary Ovarian Insufficiency (POI) | | |
|---|---|---|---|---|---|---|---|---|
| | Age | Sex | ROR | 95% CI | p | ROR | 95% CI | p |
| 2006-24 | 6-17 | F | 11.1 | 8.0-15.4 | <0.0001 | 10.0 | 7.5-13.3 | <0.0001 |
| | 18-29 | F | 18.2 | 9.4-35.3 | <0.0001 | 11.3 | 8.0-15.9 | <0.0001 |
| | 6-17 | M | 8.3 | 4.2-16.2 | <0.0001 | | | |
| 2006-15 | 6-17 | F | 7.9 | 3.7-16.6 | <0.0001 | 7.1 | 5.0-10.2 | <0.0001 |
| | 18-29 | F | 6.9 | 1.9-24.5 | 0.001 | 6.8 | 4.5-10.3 | <0.0001 |
| | 6-17 | M | 11.8 | 3.4-40.3 | <0.0001 | | | |

Table 1: Reporting odds ratios (ROR) for postural orthostatic tachycardia syndrome (POTS) and for primary ovarian insufficiency

**Comparison with Related Studies**

For POTS, these results corroborate the results of an earlier study by Chandler et al. (Drug Safety, 2017), that used partly overlapping spontaneous reports but with some key differences. Most importantly, the longer time period of 2006-2024 provides more reliable data with a large sample size despite only using the VAERS data from the USA rather than Chandler's global VigiBase data.

Arana et al. (2017) have, just like Tomljenovic, analyzed VAERS data. They used data from 2006 to 2015, looking specifically at POTS. Their focus is on the descriptive nature of the data. Because of under reporting in spontaneous reporting data, their attempt to calculate rates using the number of distributed vaccines is of little value, as explained above reading the MAH response to EMA. They also analyze POTS more formally using empirical Bayes Poisson shrinkage, which is a data mining method. That is also highly problematic though. Data mining methods are great when searching for unknown and unspecified problems, but they should not be used when a specific predefined hypothesis is being evaluated, as that would greatly reduce the statistical power and ability to conclude that a problem exist (DuMouchel 1999; Kulldorff et al., 2003).

## 5. COMMENTS ON CHAO ET AL: SURVEILLANCE OF AUTOIMMUNE CONDITIONS FOLLOWING ROUTINE USE OF QUADRIVALENT HUMAN PAPILLOMAVIRUS VACCINE (2011)

Chao et al. (2011) conducted a study to evaluate whether Gardasil, a human-papilloma virus (HPV) vaccine, may cause different forms of auto-immune diseases. This is an important question. If there is a safety problem with HPV vaccines, that is important to know, and if so, in what population groups, so risks can be adequately stated and communicated, and possibly mitigated. If there are no safety problems, that is also important to know and show, so that the public can have high confidence in the vaccine.

When evaluating the quality of a scientific study, there are four critical components to consider: the quality of the data used, the epidemiological design of the study, the statistical analysis, and the interpretation. The Chao paper has four different designs/analyses to evaluate: the main analysis, sensitivity analysis #1, sensitivity analysis #2 and the time to event data.

**Data**

The study uses data from Kaiser Permanente Southern California (KPSC) and Kaiser Permanente Northern California (KPNC). This is the best observational data available for vaccine safety studies. As the two largest components of CDC's Vaccine Safety Datalink, the world's preeminent vaccine safety surveillance system, KPSC and KPNC allows for large sample size analysis, which is important when evaluating rare but serious adverse events. The study was restricted to girls and women ages 9 to 26. While it is obviously also of interest to know the effects of the Gardasil vaccine on boys and young men, the restriction to females was a natural and logical choice considering the vaccine uptake at the time of the study.

The authors choose a limited set of auto-immune diseases to study. As a methodologist, I have no opinion on this choice. While it would have been interesting to study additional auto-immune diseases, and while the sum of the sample size from KPSC and KPNC would have made that possible, one study cannot study everything. It is more important that whatever is studied, it is studied in a scientifically sound way. In summary, KPSC and KPNC data was an excellent choice for this study, and it is hard to imagine a better choice.

**Main Analysis**

In the main analysis, the intent is to compare the frequency of selected auto-immune diseases between vaccinated and unvaccinated individuals, after chart review. This analysis has two fatal flaws.

*Missing Data Imputation*

Missing data imputation is a useful tool when some data is missing. For example, a study of vaccines and febrile seizures should adjust for age, but if age is missing for some, age may be imputed for those individuals rather than having to exclude them from the analysis. If age is missing for most individuals, one must accept that limitation and recognize that it is impossible to adjust for age rather than imputing age for almost everyone.

In the Chao study, imputations were made regarding the outcome of interest for every single individual in the unvaccinated comparator group. That is astonishing and something that I have never seen before. It is ill-advised to impute data for the main outcome of interest, and irrespectively of the variable, it is ill-advised to impute it for every individual in one of the comparator groups.

If there were financial limitations on the number of chart reviews that could be conducted, Chao et al. should instead have done chart-review of both vaccinated and unvaccinated individuals in about equal numbers. If that was not possible for some

logistical or other reasons, they should have accepted that it was impossible for them to do chart confirmed analysis comparing vaccinated and unvaccinated individuals, instead analyzing the unreviewed electronic data.

To credit the authors, they do recognize in the discussion *"that the multiple imputation approach was not a standard method for estimating background incidence rates."* I hope and trust that it never will be. Moreover, as a limitation they note that *"only the reviewed vaccinated cases were used to inform the imputation for new-onset status amongst the unvaccinated potential cases."* It is not just a limitation though, but a fatal flaw.

*Assuming Equal Chart-Confirmation Rates for Vaccinated and Unvaccinated*

To do the missing data imputation, Chao et al. assumed that the chart-confirmation rate would be the same for vaccinated and unvaccinated individuals. That is only correct if the null hypothesis is true. If the vaccine causes the adverse event, it is mathematically impossible for the two rates to be the same (unless there are no false positives or the vaccine has a causal relationship with the events that are classified as false positives). That is, this critical assumption made by the authors is only true if the vaccine does not cause the adverse event under study. If Gardasil does cause the auto-immune disease outcome of interest, the chart-confirmation rates will be different. To illustrate this, consider the following hypothetical scenario.

Let F1 and F2 be the proportion of false positives among the vaccinated and unvaccinated, respectively. Since these are false positives, they are unrelated to the vaccine, so F1=F2=F, and since there were some false positives in their data, we know that F>0. Let T1 and T2 be the proportion of true positives among the vaccinated and unvaccinated respectively. If the null hypothesis is true, then T1=T2. If the vaccine causes some of the adverse events, then T1>T2.

What does this imply for the chart confirmation rates? For the vaccinated, the chart-confirmation rate is C1=T1/(F+T1), while it is C2=T2/(F+T2) for the unvaccinated. Let x=T1-T2 be the difference in the proportions of true positives. We then have C1=(x+T2)/(F+x+T2) and C2=T2/(F+T2).

Chao et al. assume that C1=C2, that is, they assume that C1-C2=0. When is that a correct assumption? Using the above formula, if x=0, then C1=C2, so their assumption is correct when the null hypothesis of T1=T2 is true. It is also correct when F=0, that is, when there are no false positives, in which case the chart-confirmation rate is 100% among both the vaccinated and the unvaccinated. For other scenarios, when both F>0 and x>0, we have

$$\begin{aligned} C1\text{-}C2 &= (x+T2)/(F+x+T2) - T2/(F+T2) \\ &= [\,(x+T2)\,(F+T2) - T2(F+x+T2)\,] \,/\, [\,(F+T2)(F+x+T2)\,] \\ &= [\,(x*F+x*T2+T2*F+T2*T2) - (T2*F+T2*x+T2*T2)\,] \,/\, [\,(F+T2)(F+x+T2)\,] \\ &= x*F \,/\, [\,(F+T2)\,(F+x+T2)\,] \end{aligned}$$

Since F, x and T2 are all positive, both the numerator and the denominator are positive. This means C1-C2>0. This in turn means that the assumption that Chao et al. makes about equal chart confirmations rates is wrong when the vaccine causes the adverse event. It is only correct if the null hypothesis that the vaccine does not cause the adverse event is true.

For a more concrete illustration, consider the following hypothetical example. Suppose we have a cohort of 100,000 vaccinated persons and 100,000 unvaccinated. Some of them will have a false positive diagnosis due to coding errors or another disease with some overlapping symptoms. These are unrelated to the vaccine, so they occur in equal numbers of say 200 in each of the two groups. We also have true positives. Some are unrelated to the vaccine, and let's assume there are 100 of those among both the unvaccinated and vaccinated. If the vaccine causes the adverse event

of interest, there will be more true positives among the vaccinates, say 300. If we now compare the incidence rates, the chart-confirmed incidence rate is 300 per 100,000 in the vaccinated compared to 100 per 100,000 among the unvaccinated. No surprise there. If we instead compare the chart-review confirmation rates, it is 300 / (300+200) = 60 percent among the vaccinated while it is 100 / (100+200) = 33.3 percent among the unvaccinated. By modifying these numbers, we can see that the difference in the chart-review confirmation rates depends both on the number of false positives and the difference in the number of true positives.

If the vaccine increases the risk of an adverse event, then the chart-confirmation rate is lower among the unvaccinated. This means that by falsely assuming that the chart-confirmation rates are the same in the unvaccinated as in the vaccinated, that introduces a bias towards the null hypothesis, making it more difficult to detect any causal adverse reactions to the vaccine. Except for the direction, this is acknowledged by Chao et al., when they accurately state that *"If the true new-onset confirmation rate was lower in the unvaccinated population, this may have biased the IRR estimates"* (p.202). It was also subsequently acknowledged in a commentary by Chao and Jacobsen (2012), when they write that *"one way to address the concern of lead time bias is to conduct case review of potential new onset cases arising from unvaccinated subjects as well."*

In summary, Chao et al. assumed that the chart confirmations rates were equal between the vaccinated and unvaccinated, after adjusting for patient predictors such as age. This is only true when the null hypothesis is true. Hence, when evaluating whether the vaccine causes the adverse event, they assumed that the vaccine does not cause the adverse event, or in other words, when evaluating whether the null hypothesis is true or not, they are assuming that the null hypothesis is true. This is a circular argument. It is the second fatal flaw of the main analysis.

Note that the inclusion of age and other predictors in the "imputations" does not overcome the problem that they assumed equal chart confirmation rates between the unvaccinated and the vaccinated. They simply assumed that the chart confirmation rate for an unvaccinated person with certain age and other characteristics were the same as for a vaccinated person with the same age and characteristics. Nor does the inclusion of age in the imputation algorithm mean that the analysis itself was adjusted for age. In fact, it is rather perplexing that the authors would account for age and other covariates in their complex imputation algorithm but not in the actual analyses.

There are additional problems with the main analysis, such as the chart reviewers not being blinded to vaccination status and failure to adjust for age in the actual analysis, but if a boat does not float, what's the point in criticizing its sail or rudder?

**Sensitivity Analysis #2**

Sensitivity analysis #2 has the same fatal flaws as the main analysis.

**Sensitivity Analysis #1**

Since both the main analysis and sensitivity analysis #2 are fatally flawed, we are left with sensitivity analysis #1. Since this analysis does not use any chart review at all, there is no missing data imputation problems and there is no assumption about equal chart review confirmation rates. Hence, this analysis does not suffer from the two fatal flaws described above.

*Lack of Chart-Review*

While it is much better to analyze chart-confirmed data when done correctly, it is easier and cheaper to analyze unconfirmed electronic health records. If the vast majority of the events are true positives, such an analysis is almost as good as analyzing chart-confirmed data. When the proportion of true positives is low, as in the Chao article, there is both a substantial loss in power and a considerable

attenuation of the effect size towards the null. This is a limitation of sensitivity analysis #1, but it is not a fatal flaw.

*Sample Size*

Sensitivity analysis #1 used only data from KPSC. The loss in power due to the lack of chart review can be at least partly compensated by using a larger sample size. The authors had the opportunity to do so effortlessly, but did not. One of the great features of the KPSC and KPNC data is the common data model. This makes it possible to run exactly the same SAS code on both data sets with minimal additional effort, with only a few additional hours from a data manager. It is therefore surprising and incomprehensible that the authors only used KPSC data for this analysis. While not a flaw in the analysis, it is a major flaw in the design.

*Lack of Age Adjustment*

The Gardasil vaccination rate varies greatly by age, with the highest around the recommended age of around 11 to 13. The incidence rates of many of the studied auto-immune disease outcomes also varies by age. When there is age variation in both exposure and outcome, it is important to adjust the analysis for age. If not, the results will be biased. The bias can go in either direction. Age is an available variable in both the KPSC and KPNC data, and from a statistical perspective, it is not difficult to adjust the analyses for age. Unless they simply failed to mention it, I do not understand why such an adjustment was not made, especially since Chao and Jacobsen (2012) subsequently pointed out that *"while there are few well established risk factors for autoimmune conditions, factors such as age, sex, race, genetic predisposition and certain environmental exposure should be considered in observational studies evaluating autoimmune safety signals."*

Note that an age-adjustment is very different from age stratification. In age-stratified analyses, a separate statistical analysis is done for each age group, greatly reducing the statistical power and ability to detect potential problems. With age-adjustment,

a single statistical analysis is done for all age groups combined, which maintains the statistical power, but any age-related bias is removed.

Since only girls and women were included in the study, there was no need to adjust for gender. There may have been other important variables to adjust for, but age was most likely the most important one.

In summary, the authors did not utilize as large a sample size as they easily could have, and they failed to adjust for age, which are flaws of sensitivity analysis #1.

**Time to Event Data**

A powerful way to analyze vaccine safety data is to look at the time between the day of vaccination and the day of the event. If unrelated to the vaccine, adverse events should be evenly distributed over the days, weeks and months after the vaccination day. As a contrast, if the vaccine causes an adverse reaction, those will sometimes cluster a certain number of days after the vaccination. For example, for one-year-old children, there is an increased risk of seizures 7 to 11 days after the MMRV vaccine (Klein et al. 2010). A great advantage of this study design is that it is self-controlled, in that the control is a different time period from the same period rather than a different individual or group of individuals.

Chao et al. collected the relevant data to conduct a self-control analysis, including the time of the vaccine, the dose, the timing of the outcome and the length of follow-up. The latter is critical. If the majority of recorded adverse events occurred within a month of vaccination, that is expected if most people only have one month of follow-up, but it would be highly unusual and suspect of most patients had a year of follow-up time.

In Figure 2, Chao et al provide the timing vaccines and adverse events, as well as the dose number. Since the follow-up time is not provided, the information in the graph

is useless. To illustrate, let's consider the top graph for diabetes. At least one patient had at least 600 days of follow-up, since he or she received a second dose around that time. If all the other patients also had 600 days follow-up, there is a clear and obvious cluster of cases between zero and 300 days after the first dose of the vaccine, since only one out of 15 cases occurred after this time period. On the other hand, if most patients had a follow-up of less than 300 days, it is just natural that most recorded adverse events also occurred within 300 days of vaccination. The fact that Figure 2 does not include any information on follow-up time, makes it useless for evaluating adverse events. It is very surprising that the authors did not include that information.

There are a variety of ways to conduct a formal statistical analysis of self-control data, the simplest being the comparison of two equal length risk and control windows for each patient, all the way to more sophisticated temporal scan statistics with variable window lengths. Surprisingly, Chao did not do any formal statistical analysis of these data. Hence, it is impossible to know whether the data has information about potential adverse reactions, or if they are just random noise.

**Interpretation**

In the discussion, Chao et al. asserts that *"there was no clear evidence of safety signal for autoimmune conditions following vaccination with HPV4 in this study."* This is technically truthful, just as an article on butterflies would not have any such evidence, but the sentence is highly misleading. Because of the fatal flaws, no useful information at all can be derived from the main analysis or from sensitivity analysis #2.

A well-designed vaccine safety study will either find a problem or provide reassurance that the vaccine is safe with respect to the outcome studied. Because of the fatal flaws, the main analysis does not provide any reassurance regarding the safety of the Gardasil vaccine. The same is true for the time to event data in Figure 2, but for

different reasons, that critical information was withheld and there were no formal statistical analyses conducted.

The only analysis not completely devoid of information is sensitivity analysis #1, but its statistical power is low due to only using KPSC data and due to high false positive rates, and the analyses are likely biased due to the failure to adjust for age.

The only potentially useful conclusion one could make from the Chao et al. study is that some of the studied outcomes warrant a thorough evaluation using a properly designed and appropriately powered study; most urgently systemic lupus erythematosus and Hashimoto's disease. This in addition to other auto-immune conditions not studied by Chao et al., such as POTS and POI

My recommendations are that the authors:

(i) retract the paper

(ii) redo sensitivity analysis #1 using both KPSC and KPNC and with data up until present time, while adjusting at least for age

(iii) conduct a proper self-control analysis of the time to even data, as some of the authors have done for other vaccines and outcomes

(iv) conduct chart review on all potential cases of both vaccinated and unvaccinated individuals, followed by an age-adjusted analysis. If resources are limited, it is better to evaluate a few outcomes with well powered studies, rather than scatter resources and sample size over many outcomes generating a large number of underpowered studies.

## 6. CONCLUSIONS

### Proper Evaluation of Vaccine Safety

In my expert opinion, Merck has not done an appropriate job evaluating potential safety problems with the Gardasil vaccine. Vaccines are important for public health,

and timely and appropriate investigation of potential safety concerns are important to maintain a high confidence and trust in not only the specific vaccine in questions, but also in vaccines in general. By responding to safety concerns about the Gardasil vaccine with reports and studies that are fatally flawed, Merck has done a great disservice to public health, irrespectively of whether there is a problem with the vaccine or not.

When there is an adverse reaction to a vaccine, that is important to quickly detect, evaluate, establish and quantify. It does not necessarily mean that the vaccine should be withdrawn from the market, but it is important to know the type of adverse reaction and the population at risk. For example, the RotaTeq vaccine against rotavirus is still on the market even though it causes intussusception in very small number of children, but it is critically important to know about this problem so that there is no delay in diagnosis and treatment when it happens. As another example, the measles-mumps-rubella-varicella (MMRV) vaccine is still on the market even though it can cause febrile seizures, although it is not used for the first dose at 12-18 months of age.

When there is no problem with a vaccine, it is equally important to properly evaluate any potential adverse events. If a suspected problem is allowed to linger, without properly conducted studies to refute it, the public rightly loses trust in both the vaccine in question and in other vaccines.

**Safety of the Gardasil Vaccine**

The Chao paper contributed nothing regarding potential adverse reactions to the Gardasil vaccine, in either direction. While the MAH report provides some descriptive data, proper analyses were not included, and the limited analyses were flawed. Hence, neither the Chao paper nor the MAH report can be used to either confirm or refute a safety problem with the Gardasil vaccine.

Using VAERS data, the reporting odds ratios presented by Dr. Tomljenovic for POTS and POI are concerning. With spontaneous reporting data, one should be cautious about interpreting small reporting odds ratios, but for POTS and POI they are so large that they are unlikely to be due to bias from a lack of adjustment for age or other demographic or clinical variables. It is more difficult to judge potential reporting bias.

VAERS data are often misused, sometimes by assuming that all reported adverse events are due to the vaccine and sometimes by calculating rates by dividing the number of reported adverse events with the number of distributed vaccines, as Merck did. Dr. Tomljenovic analyzed the data properly using reporting odds ratios.

Safety signals from spontaneous reporting systems should always be followed-up and evaluated further. As in all research, it is important to conduct multiple studies using different data sets and different designs, before any definite conclusions can be made about causality. When doing so, it is critical to (i) properly define the relevant disease outcome, and in many cases, do careful chart review to minimize the number of false positives and false negatives, (ii) properly define the population at risk, (iii) evaluate the outcome in an unbiased manner between exposed and unexposed and account for potential confounders, and (iv) have adequate sample size for rare but serious adverse events.

In the study of POTS and POI after the Gardasil vaccine, the first item is especially critical, since there are no well-established and adequate ICD-10 codes for these syndromes. In the Japanese study by Fukushima et al. (2022), there were a modest 40% excess among the vaccinated when looking at people with only a few symptoms, but a three-fold excess when looking at people with multiple simultaneous symptoms. While they did not conduct formal statistical analyses of their data, it illustrates the need to carefully define the outcome so as not to attenuate the results by including a wider range of conditions unrelated to the vaccine. For spontaneous reporting data,

such as VAERS, this is less of a problem, but it is an issue when using electronic health data. For outcomes that are well described by ICD-10 codes, electronic health data can be analyzed without chart reviews, but to study POTS and POI after vaccination, well defined syndromes and chart review are critical.

The second key aspect is the proper definition of the population at risk, both in terms of demographics and the risk period. The well-established excess risk of febrile seizures after the MMRV vaccine is a perfect example of this issue. (Klein et al., 2010) The risk is only present for the first dose at ages 12-18 months, so an analysis of the second dose at age 4-6 years would give the impression that the vaccine is safe (which it is for that age group), while a combined analysis would attenuate the results and possibly miss the problem for the younger children. The excess risk of febrile seizures occurs 7-10 days after vaccination, but if one analyses a longer time period, that excess risk will be attenuated and undetectable due to added random noise to the data. At the same time, the opposite can be true. The diagnosis of a less acute condition may be spread over many months or years, making it impossible to detect if the exposure window is too short or if a self-control analysis is conducted.

While double-blind randomized controlled trials are ideal to avoid bias between the exposed and unexposed, their sample size is often too small to evaluate rare but serious adverse events, making properly designed and conducted observational studies critical. A major strength of the latter is their large sample size and the ability to easily identify potential patients with the conditions of interest for detailed chart review.

In conclusion, after concerns about Gardasil induced auto-immune diseases were raised, there were many options for conducting high quality observational studies to thoroughly evaluate those concerns. It is unfortunate that Merck declined to do so.

Dated: September 9, 2024.

## 7. REFERENCES

Arana J, Mba-Jonas A, Jankosky C, Lewis P, Moro PL, Shimabukuro TT, Cano M. Reports of postural orthostatic tachycardia syndrome after human papillomavirus vaccination in the vaccine adverse event reporting system. Journal of adolescent health. 2017 Nov 1;61(5):577-82.

Chandler RE, Juhlin K, Fransson J, Caster O, Edwards IR, Norén GN. Current safety concerns with human papillomavirus vaccine: a cluster analysis of reports in VigiBase®. Drug safety. 2017 Jan;40:81-90.

Chao C, Klein NP, Velicer CM, Sy LS, Slezak JM, Takhar H, Ackerson B, Cheetham TC, Hansen J, Deosaransingh K, Emery M. Surveillance of autoimmune conditions following routine use of quadrivalent human papillomavirus vaccine. Journal of internal medicine. 2012 Feb;271(2):193-203.

de Mos M, De Bruijn AG, Huygen FJ, Dieleman JP, Stricker BC, Sturkenboom MC. The incidence of complex regional pain syndrome: a population-based study. Pain. 2007 May 1;129(1-2):12-20.

DuMouchel W. Statistical Practice. The American Statistician. 1999 Aug;53(3).

Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiology and drug safety. 2001 Oct;10(6):483-6.

Fukushima W, Hara M, Kitamura Y, Shibata M, Ugawa Y, Hirata K, Oka A, Miyamoto S, Kusunoki S, Kuwabara S, Hashimoto S. A nationwide epidemiological survey of adolescent patients with diverse symptoms similar to those following

human papillomavirus vaccination: background prevalence and incidence for considering vaccine safety in Japan. Journal of epidemiology. 2022 Jan 5;32(1):34-43.

Klein NP, Fireman B, Yih WK, Lewis E, Kulldorff M, Ray P, Baxter R, Hambidge S, Nordin J, Naleway A, Belongia EA. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. Pediatrics. 2010 Jul 1;126(1):e1-8.

Kulldorff M, Fang Z, Walsh SJ. A tree-based scan statistic for database disease surveillance. Biometrics. 2003 Jun;59(2):323-31.

Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. A maximized sequential probability ratio test for drug and vaccine safety surveillance. Sequential analysis. 2011 Jan 21;30(1):58-78.

Li R, Weintraub E, McNeil MM, Kulldorff M, Lewis EM, Nelson J, Xu S, Qian L, Klein NP, Destefano F. Meningococcal conjugate vaccine safety surveillance in the Vaccine Safety Datalink using a tree-temporal scan data mining method. Pharmacoepidemiology and drug safety. 2018 Apr;27(4):391-7.

Merck, Gardasil, Gardasil 9 and Silgard vaccine, Response to PRAC article 20 of regulation (EC) no 726/2004, July 2015.

Rothman KJ, Lanes S, Sacks ST. The reporting odds ratio and its advantages over the proportional reporting ratio. Pharmacoepidemiology and drug safety. 2004 Aug;13(8):519-23.

Sandroni P, Benrud-Larson LM, McClelland RL, Low PA. Complex regional pain syndrome type I: incidence and prevalence in Olmsted county, a population-based study. PAIN®. 2003 May 1;103(1-2):199-207.

Tomljenovic L, Disproportionality analysis of the U.S. VAERS data shows a signal of disproportionate reporting of POTS, CFS, CRPS, peripheral neuropathy, primary ovarian failure and menstrual abnormalities associated with Merck's Gardasil vaccines, September 6, 2024.

Yih WK, Kulldorff M, Fireman BH, Shui IM, Lewis EM, Klein NP, Baggs J, Weintraub ES, Belongia EA, Naleway A, Gee J. Active surveillance for adverse events: the experience of the Vaccine Safety Datalink project. Pediatrics. 2011 May 1;127(Supplement_1):S54-64.

Yih WK, Maro JC, Nguyen M, Baker MA, Balsbaugh C, Cole DV, Dashevsky I, Mba-Jonas A, Kulldorff M. Assessment of quadrivalent human papillomavirus vaccine safety using the self-controlled tree-temporal scan statistic signal-detection method in the Sentinel System. American journal of epidemiology. 2018 Jun 1;187(6):1269-76.